

International Journal of Population Data Science

Journal Website: www.ijpds.org



The UK Synthetic Data Community Group – The VSTAR Framework for Responsible Synthetic Data

Lewis Hotchkiss¹, Emma Squires¹, Simon Thompson¹, Emily Oliver², Sophie McCall³, Cristina Magder⁴, Timothy Rittman⁵, John Gallacher⁶, Anmol Arora⁷, and Fiona Lugg-Widger⁸

¹Swansea University, Swansea, United Kingdom; Dementias Platform UK, Swansea, United Kingdom

²Administrative Data Research UK, London, United Kingdom

³Research Data Scotland, Edinburgh, United Kingdom

⁴UK Data Service, Essex, United Kingdom

⁵University of Cambridge, Cambridge, United Kingdom

⁶University of Oxford, Oxford, United Kingdom

⁷University College London, London, United Kingdom

⁸Cardiff University, Cardiff, United Kingdom

Synthetic data is increasingly recognised as a powerful tool for accelerating research, training, and innovation within Trusted Research Environments (TREs), but its safe and responsible use requires clear governance standards and shared best practices. To address this need, the UK Synthetic Data Community Group (SDCG) was established with DARE UK funding to convene leaders across the synthetic data landscape and build national consensus on responsible synthetic data development and release. The UK SDCG engaged more than 140 stakeholders from over 35 organisations across academia, industry, government, and the public. Through a series of workshops involving researchers, domain experts, data custodians, and public contributors, we gathered perspectives on opportunities, risks, and expectations surrounding the use of synthetic data in secure environments. Discussions focused on governance challenges, quality and utility requirements, disclosure

risk considerations, and transparency needs for users and data owners. These engagements informed the development of the VSTAR Framework, which contains five core principles for responsible synthetic data practice: Valuable, Safe, Transparent, Accessible, and Representative. Together, these principles provide a structured, practical foundation for TREs and data custodians to guide the generation, evaluation, and dissemination of synthetic datasets. The framework aims to promote consistency across infrastructures, strengthen trust and accountability, and support the adoption of synthetic data as a complementary tool that enhances secure access to real-world data. The VSTAR Framework represents a national step toward coherent governance for synthetic data, offering a scalable and community-driven model to support privacy-preserving innovation across the UK's data research ecosystem.

